

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-222339

(43)Date of publication of application : 11.08.2000

(51)Int.Cl.

G06F 13/14
G06F 3/06
G06F 13/00
G06F 13/10

(21)Application number : 11-024648

(71)Applicant : HITACHI LTD

(22)Date of filing : 02.02.1999

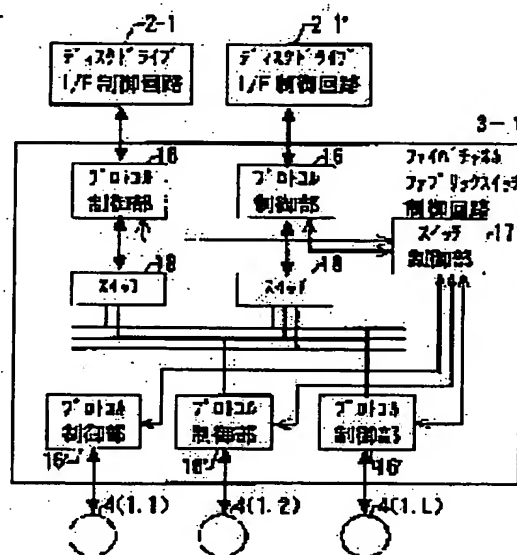
(72)Inventor : ARIGA KAZUHISA

(54) DISK SUB-SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To connect plural disk drives with a disk drive interface circuit without scarifying transmitting performance by using a fiber channel fabric topology for reducing the number of connection lines by using a fiber channel interface being a serial interface, and for realizing switch connection.

SOLUTION: A fiber channel fabric switch control circuit 3 is provided between a disk drive 4 and a disk drive interface control circuit 2, and protocol control part 16 is provided between a switch 18 in the fabric switch circuit 3 and the disk drive.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-222339

(P2000-222339A)

(43) 公開日 平成12年8月11日 (2000.8.11)

(51) Int. Cl.	識別記号	F I	テマート (参考)
G 0 6 F 13/14	3 1 0	G 0 6 F 13/14	3 1 0 F 5 B 0 1 4
3/06	3 0 1	3/06	3 0 1 A 5 B 0 6 5
	3 0 5		3 0 5 C 5 B 0 8 3
	5 4 0		5 4 0
13/00	3 0 1	13/00	3 0 1 D

審査請求 未請求 請求項の数 5 O L (全 10 頁) 最終頁に続く

(21) 出願番号 特願平11-24648

(22) 出願日 平成11年2月2日 (1999.2.2)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 有賀 和久

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 100068504

弁理士 小川 勝男

Fターム (参考) 5B014 EA02 EA04 EB05 HA09 HA12

5B065 BA01 CA11 CA19 CA30 CE12

EA25 ZA11

5B083 AA08 BB01 BB03 CC02 CD13

EE08 EF11

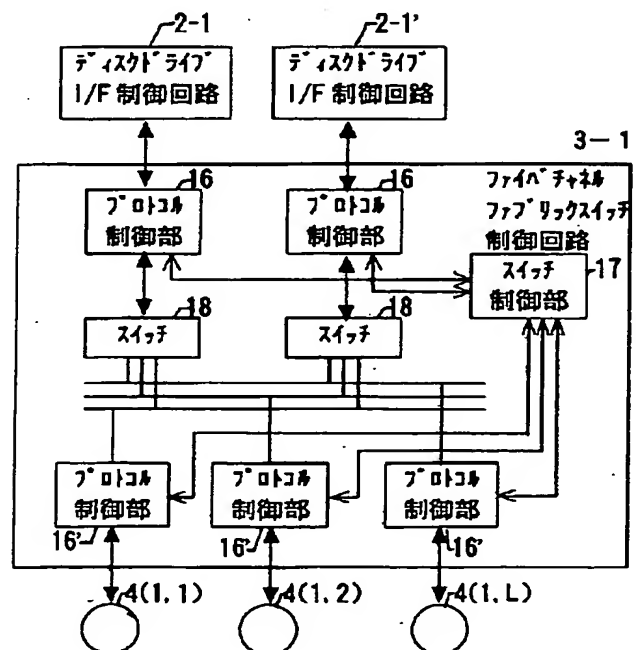
(54) 【発明の名称】 ディスクサブシステム

(57) 【要約】

【課題】 ディスク制御装置とディスクドライブの接続においては、SCSIを用いたインタフェースが主流であるが、ディスクドライブが増加した場合に1本のインタフェースで1対1の接続を行う場合、現状のファイバチャネルを用いたディスクドライブでは、スイッチ接続が出来ない形式となっているので多数のインタフェースが必要となり、実装面で困難が生ずる。

【解決手段】 ディスクドライブ4とディスクドライブインタフェース制御回路2との間にファイバチャネル・ファブリック・スイッチ制御回路3を設け、このファイバチャネル・ファブリック・スイッチ回路3内のスイッチ18とディスクドライブとの間にプロトコル制御部16を設ける。

図 3



【特許請求の範囲】

【請求項1】データを記憶する複数のディスクドライブと、このディスクドライブ及びホストコンピュータからのデータの入出力を制御するディスクアレイ制御部とを有し、このディスクアレイ制御部と前記ディスクドライブとをファイバチャネルにて接続したディスクサブシステムにおいて、

前記ディスクアレイ制御部と前記ディスクドライブとをスイッチ接続したディスクアレイシステム。

【請求項2】データを記憶する複数のディスクドライブと、このディスクドライブ及びホストコンピュータからのデータの入出力を制御するディスクアレイ制御部とを有するディスクサブシステムにおいて、

前記ディスクドライブと前記ディスクアレイ制御部との間にスイッチとこのスイッチの切換え制御をするスイッチ制御部を設け、前記スイッチと前記ディスクドライブとの間、及び/または前記ディスクアレイ制御部と前記スイッチとの間にプロトコル制御部を設けたディスクサブシステム。

【請求項3】前記ディスクアレイ制御部と前記スイッチとの間、及び前記スイッチと前記ディスクドライブとの間とはファイバチャネルを用いて接続したものであり、前記スイッチはファイバチャネルファブリックスイッチである請求項2に記載のディスクサブシステム。

【請求項4】ホストコンピュータからのデータの入出力を制御するホストインタフェース制御部と、このホストインタフェース制御部で受けたデータを一時的に格納するキャッシュメモリと、前記データにパリティデータを付加するパリティデータ生成部と、前記データ及び前記パリティデータを記憶する複数のディスクドライブと、このディスクドライブに前記データを書き込むディスクドライブインターフェイスとからなるディスクアレイ制御部とを有するディスクサブシステムにおいて、前記ディスクドライブインターフェイスにプロトコル制御部とスイッチを設け、前記複数のディスクドライブをスイッチ接続したディスクサブシステム。

【請求項5】ホストコンピュータからのデータの入出力を制御するホストインタフェース制御部と、このホストインタフェース制御部で受けたデータを一時的に格納するキャッシュメモリと、前記データにパリティデータを付加するパリティデータ生成部と、前記データ及び前記パリティデータを記憶する複数のディスクドライブと、このディスクドライブに前記データを書き込むディスクドライブインターフェイスとからなるディスクアレイ制御部とを有するディスクサブシステムにおいて、前記ディスクアレイ制御部と前記ディスクドライブとの間をファイバチャネルを用いて接続し、前記ディスクアレイ部と前記ディスクドライブとの間に、アクセス対象となる前記ディスクドライブのID番号検出及びファイバチャネル・プロトコルの制御を行い前記ディスクドラ

イブインタフェースと接続される第一のプロトコル制御部と、各ディスクドライブのID番号を記憶しておりID番号によりスイッチを設定するスイッチ制御部と、前記ディスクドライブにこのID番号を割り付ける前記ディスクドライブ4と接続される第二のプロトコル制御部とを備えたファブリックスイッチを設けたディスクサブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ディスクサブシステム、ディスクアレイ、ディスクドライブを内蔵した計算機等の電子機器に関し、特にアレイディスクをファブリックスイッチ接続し高速転送を可能とする技術に関する。

【0002】

【従来の技術】一般に、ディスクアレイにおいてディスク制御装置と複数のディスクドライブとを接続する場合には、特開平10-171746号公報に記載のようにSCSIインタフェース、若しくはファイバチャネル・アービトレイテッドループ・トポロジが利用されている。

【0003】SCSIインタフェースは、同一線路上にデータを時分割して転送する方式をとっており、イニシエータに対するアクセスは、1伝送路上に1時刻あたり1対1の通信を行う方式である。

【0004】ファイバチャネル・アービトレイテッドループ・トポロジでは、SCSIインタフェースに対して、シリアルインタフェースによりループ状にイニシエータ、ディスクドライブを接続することができ、フレームに分割されたデータを時分割して転送し、同時に多数デバイスの通信が行え、接続可能ディスクドライブ数も126と拡張できる。

【0005】

【発明が解決しようとする課題】今後ディスクドライブの小型化・高密度化により、より多くのディスクドライブを使用することが可能となると考えられる。

【0006】SCSIインタフェースは、1伝送路上に1時刻あたり1対1の通信を行う方式であるため同時に多数のイニシエータとディスクドライブの通信ができない。また、接続可能なディスクドライブの数も7から15台と少ない。そのためSCSIを用いたインタフェースでドライブが増加した場合に1本のインタフェースで1対1の接続を行おうとすると、多数のインタフェースが必要となり、実装面で困難が生ずる。また、1つの制御回路で接続可能なディスクドライブの数が少ないため、多数の制御回路を使用する必要が生じる。

【0007】一方、ファイバチャネルを使用した場合、ディスクドライブはプロトコルが制御装置とは異なるためにスイッチ接続が出来ず、多数のディスクドライブが同一ループを共有するファイバチャネル・アービトレイ

テッドループを用いてループ接続とせざるを得なかった。そのため、同一ループに接続されるディスクドライブ数を増加すると、ディスクドライブのデータ転送速度がループの最大データ転送速度よりも大きくなり、結果的にループの最大データ転送速度以上の効率では転送が行えなくなりSCSIインタフェースと同程度のデータ転送速度でしか接続できなかった。

【0008】

【課題を解決するための手段】上記課題を解決するため本発明では、ディスクドライブと制御装置とをスイッチ接続を可能とするため、プロトコル制御部をファイバチャネル・ファブリック・スイッチとディスクドライブとの間に設ける。

【0009】

【発明の実施の形態】以下、図面を用いて本発明を適用した外部記憶装置（ディスクサブシステム）の実施例を説明する。図1は全体図である。

【0010】図に示す外部記憶装置において、N個のディスクアレイ制御回路（制御部）（1-1）～（1-N）（途中の1-2等は省略、以下同じ）は、上位側はホストコンピュータ（図示せず）に接続され、下位側はM個のディスクドライブインターフェイス（ディスクドライブI/F）制御回路（2-1）～（2-M）を備えている。ディスクアレイ制御回路のハード構成の詳細は後述する。M個のファイバチャネル・ファブリック・スイッチ制御回路（3-1）～（3-M）は、ファイバチャネル・インタフェース5によってディスクドライブを制御するディスクドライブインターフェイス（I/F）制御回路（2-1）～（2-M）にそれぞれ接続されている。そして一つのファイバチャネル・ファブリック・スイッチ制御回路に対してL個、計M×L個のディスクドライブ（4（1, 1）～4（M, L））は、ファイバチャネル・インタフェース6によって、ファイバチャネル・ファブリック・スイッチ制御回路（3-1）～（3-M）と接続されている。

【0011】また、各ディスクドライブインターフェイス制御回路（2-1）～（2-M）、及びデータを格納しておくディスクドライブ4（1, 1）～4（M, L）はそれぞれ個別の識別子（ID番号）をもつ。ファイバチャネル・ファブリック・スイッチ制御回路（3-1）～（3-M）は、ディスクドライブインターフェイス制御回路（2-1）～（2-M）から接続するディスクドライブのID番号を受け取り、対応するディスクドライブインターフェイス制御回路（2-1）～（2-M）とディスクドライブ4（1, 1）～4（M, L）の1対1の接続を確立する。

【0012】図2にディスクアレイ制御回路（1-1）～（1-N）のハードウェア構成を示す。上位ホストコンピュータ（図示せず）から転送されるデータは、ホストインタフェース制御部7により制御されキャッシュメ

モリ8に一時格納されると共にパリティデータ生成部9によりパリティデータを付加され、データブロックとパリティデータブロックとに分解（全体でM個）される。これらのデータ及びパリティのブロックは、それぞれ対応するインタフェースであるディスクドライブインターフェイス制御回路（2-1）～（2-M）によりディスクドライブグループ（図示せず）に格納される。

【0013】上位ホストコンピュータにデータを転送する場合は、転送するデータがキャッシュメモリ8に存在する場合には、そのデータをホストインタフェース制御部7が上位ホストコンピュータに転送する。転送するデータがキャッシュメモリ8に存在しない場合には、ディスクドライブインターフェイス制御回路（2-1）～（2-M）がディスクドライブグループより分解されたデータを読み出し、パリティデータ生成部9で分解されたデータを結合した後にキャッシュメモリ8に一時格納するとともにホストインタフェース制御部7が上位ホストコンピュータに転送する。

【0014】なお、以上の例はRAIDを用いた場合のデータ格納方法であり、RAID方式を用いずにデータを格納することも当然可能である。その場合にはパリティデータ生成部9が存在せずに上位ホストコンピュータ（図示せず）から転送されるデータは、ホストインタフェース制御部7によりキャッシュメモリ8に一時格納されると共にディスクドライブグループ内の何れかのディスクドライブに格納され、ミラー方式の場合には、複数のディスクドライブに同一のデータを複数格納する。読み出す際にもディスクドライブからデータを読み出し、キャッシュメモリ8に一時格納するとともにホストインタフェース制御部7が上位ホストコンピュータに転送する。

【0015】以下の例もRAIDを使用したディスクサブシステムについて説明するが、RAIDを用いた場合に限らないことはもちろんである。

【0016】図3にファイバチャネル・ファブリック・スイッチ制御回路（3-1）～（3-M）のハードウェア構成を示す。ディスクドライブインターフェイス制御回路（2-1）と接続されるプロトコル制御部16（第一のプロトコル制御部）は、アクセス対象となるディスクドライブ4（1, 1）～4（1, L）のID番号検出及びファイバチャネル・プロトコルの制御を行う。ディスクドライブ4（1, 1）～4（1, L）と接続されるプロトコル制御部16'（第二のプロトコル制御部）はディスクドライブ4（1, 1）～4（1, L）にID番号を割り付け、スイッチ制御部17に担当するディスクドライブ4（1, 1）～4（1, L）のID番号を報告する。スイッチ制御部17は、各ディスクドライブ4（1, 1）～4（1, L）のID番号を記憶しており、ディスクドライブインターフェイス制御回路（2-1）～（2-M）より受領したID番号によりスイッチ18を

設定し、1対1の接続を確立する。

【0017】尚、プロトコル制御はプロトコル制御部16'側で行うように設定してもよいし、ホストコンピュータからのデータ転送時とホストコンピュータへデータ転送時とや、通常データ転送とディスク障害時のデータ移送とでプロトコル制御部16とプロトコル制御部16'とを切り換えるように設定してもよい。

【0018】また、プロトコル制御部16或いはプロトコル制御部16'の何れか一方のみとし、プロトコル制御部16の代わりにID番号検出手段を設ける、或いはプロトコル制御部16'の代わりにID番号割り付け手段を設けてもよい。

【0019】また、ファイバチャネル・ファブリック・スイッチを独立した装置としてではなく、ディスクドライブインタフェース制御回路(2-1)~(2-M)内にプロトコル制御部とスイッチとを設け、直にディスクドライブ4(1,1)~4(1,L)と接続するようにしてもよい。

【0020】図4にファイバチャネルファブリックスイッチ制御回路(3-1)~(3-M)の動作を示す。

【0021】ディスクアレイ制御回路(1-1)は、M個に分解されたデータをディスクドライブグループ(10-1)に格納する。この際、ディスクアレイ制御回路(1-1)のM個のディスクドライブインタフェース制御回路(2-1)~(2-M)は、ファイバチャネル・ファブリック・スイッチ制御回路(3-1)~(3-M)に対し、ディスクドライブグループ(10-1)に属するディスクドライブのID番号を送信し、スイッチの確立を行う。ファイバチャネル・ファブリック・スイッチ制御回路(3-1)~(3-M)内のプロトコル制御部16(図3参照)は、ID番号を検出し、スイッチ制御部17にスイッチ接続の切替を要求する。そしてディスクドライブに合わせたプロトコル制御を行う。スイッチ制御部17(図3参照)はスイッチ18(図3参照)を接続要求もとのディスクアレイ制御回路(1-1)と接続要求先のディスクドライブグループ(10-1)に属するディスクドライブ4とを接続するよう切り替える。

【0022】このとき、ディスクアレイ制御回路(1-1)は、ディスクドライブグループ(10-1)とファイバチャネル・ファブリック・スイッチ制御回路(3-1)~(3-M)を介して1対1で対応しているのので、他のディスクアレイ制御回路(1-N)と他のディスクドライブグループ(10-2)は独立して他のデータ転送を行うことが出来る。つまり、ディスクアレイ制御回路(1-N)がディスクドライブグループ(10-L)に対する接続の確立を行っても、ディスクアレイ制御回路(1-1)とディスクドライブグループ(10-1)及びディスクアレイ制御回路(1-N)とディスクドライブグループ(10-L)との接続は互いに独立して動

作することができるので、それぞれのディスクアレイ制御回路及びディスクドライブ間で可能となる最高のデータ転送速度でデータ転送を行うことができる。

【0023】尚、詳細は説明しないが、スイッチ制御部17は上記のスイッチ切換えを行うと共に、データ読み書きの際にディスクドライブが既に読み書きを出来る状態になったという信号を受けてスイッチ18の接続切換えを行うことで転送時間を有効に最大限確保することができる。

【0024】図5に本発明の拡張された実施例を示す。先に示した実施例において、ファイバチャネル・ファブリック・スイッチ制御回路3のプロトコル制御部16とディスクドライブ4とを1対1で対応させて接続していた部分を、プロトコル制御部16からファイバチャネル・アービトレイテッド・ループ制御回路11を介して複数のディスクドライブ4をループ接続するしている。この様に接続することで、安価なディスクドライブ4を多数の接続することで大容量のディスクドライブを備えた場合と同等な性能にできる。この場合でも、全てのディスクドライブがループ接続となる訳ではなく、見かけ上はファイバチャネル・アービトレイテッド・ループ制御回路11と多数のディスクドライブ4で一つのディスクドライブ4であるので、アクセス性能は低下することがない。

【0025】また図示はしないが、ディスクドライブのアクセス速度に対し、ファイバチャネルインタフェースの最大データ転送速度に充分余裕がある場合には、複数のディスクドライブ4をファイバチャネル・アービトレイテッド・ループ制御回路11に接続し、複数のディスクドライブを同一ループ内に接続し、ファイバチャネルの最大転送レートを複数のディスクドライブ4で共有することで、アクセス性能を低下させることなくディスクドライブ4を増加させることも可能である。

【0026】図6に図5に示した実施例に用いるアービトレイテッドループ制御回路11のハードウェア構成図を示す。

【0027】アービトレイテッドループ制御回路11は、ループバイパス回路13と複数のディスクドライブ接続ポート12、及びファブリックスイッチ接続ポート15からなる。ディスクドライブ4からはループバイパス回路切替信号14が出力され、ディスクドライブ障害時にはポートをバイパスさせ、ループを切断することなく、他の動作しているディスクドライブへ影響を与えずにディスクドライブの取り外し、追加を行うことを可能とする。

【0028】図7に本発明の他の拡張された実施例を示す。

【0029】本実施例は、各ファイバチャネル・ファブリック・スイッチ制御回路(3-1)~(3-M)に接続されるスベアディスク制御回路19と、このスベアデ

ィスク制御回路19に接続される複数のスペアディスクドライブ(4-a), (4-b)を備えている。そしてファイバチャネル・ファブリック・スイッチ制御回路3内では、故障したディスクドライブ4を含むディスクドライブグループ(図では4(1, 2)のディスクドライブグループ)と接続しているプロトコル制御部16'

(図3参照)は、スイッチ18を介してスペアディスク制御回路19と接続しているプロトコル制御部16'に接続される。何れかのディスクドライブが障害を起こした場合、ディスクアレイ制御回路(1-1)~(1-N)は、スペアディスクドライブ(4-a)または(4-b)にデータの再構築を行う。

【0030】特定のディスクドライブ4にエラーが多発し故障のおそれが出た場合には、エラーが多発するディスクドライブ4のデータをスペアディスクドライブ(4-a)または(4-b)に移管させ再構築を行う。ディスクドライブ4が完全に破損してしまいデータの移管が不可能な場合には、破損したディスクドライブ4のディスクドライブグループのデータを用いて、図2に示したキャッシュメモリ8とパリティデータ生成部9にて破損データを再生しスペアディスクドライブ(4-a)または(4-b)に書き込む。

【0031】或いは、スペアディスク制御回路19が独立して行うようにしてもよい。そのためこのスペアディスク制御回路19内にキャッシュメモリやパリティデータ再生部を備える。そして、ディスクドライブ4が完全に破損した場合には、残りのディスクドライブグループのデータをスペアディスク制御回路19で読み込み、破損データを再生してスペアディスクドライブ(4-a)または(4-b)に書き込むようにする。

【0032】そのため、故障したディスクドライブ4或いは故障箇所を修復するためにパリティデータを含め分割されたデータを記憶した各ディスクドライブからスペアディスク制御回路19へのアクセスと、ディスクアレイ制御回路(1-1)~(1-N)を介して行うディスクドライブ4(図では4(1, 1)及び4(1, L)のディスクドライブグループ)とホストコンピュータからのデータアクセスとが独立して動作可能となることで、ホストコンピュータのデータアクセスに影響を与えずにデータの再構築を行うことを可能とする。

【0033】また、障害ディスクドライブが正常ディスクドライブと取り替えられた場合も同様にして、スペアディスク制御回路15がファイバチャネル・ファブリック・スイッチ制御回路(3-1)~(3-M)に対し、スペアディスクドライブ(4-a), (4-b)と、障害ディスクドライブから取り替えられた正常ディスク

ドライブと1対1の接続を行い、ディスクアレイ制御回路(1-1)~(1-N)とディスクドライブグループ(10-1)~(10-L)(図4参照)とのアクセスを妨げることなく、独立してデータのコピーを行うことで、ホストコンピュータからのアクセスにまったく影響なく障害ディスクドライブの復旧を行うことができる。

【0034】

【発明の効果】本発明により、シリアルインタフェースであるファイバチャネルインタフェースを用い接続線数を減少させ、さらにスイッチ接続を可能とするファイバチャネル・ファブリック・トポロジを用いることでディスクドライブインタフェース回路に多数ディスクドライブを伝送性能を犠牲にすることなく接続することが可能となる。また、各制御装置、ディスクドライブグループ毎に接続を動的に切り替えることで、少数のディスクドライブ制御回路で多数のディスクドライブを制御することができる。更に、ディスクドライブ障害時のデータ移行をディスクドライブインタフェース制御回路とディスクドライブのデータ転送と独立して行うことでシステムの信頼性を向上させることができる。

【図面の簡単な説明】

【図1】本発明実施例の全体図である。

【図2】ディスクアレイ制御回路の詳細図である。

【図3】ファイバチャネルファブリックスイッチ制御回路の詳細図である。

【図4】ファイバチャネルファブリックスイッチの接続図である。

【図5】ファイバチャネルファブリックスイッチとアービトレイテッドループの接続図である。

【図6】ファイバチャネルアービトレイテッドループ制御回路の詳細図である。

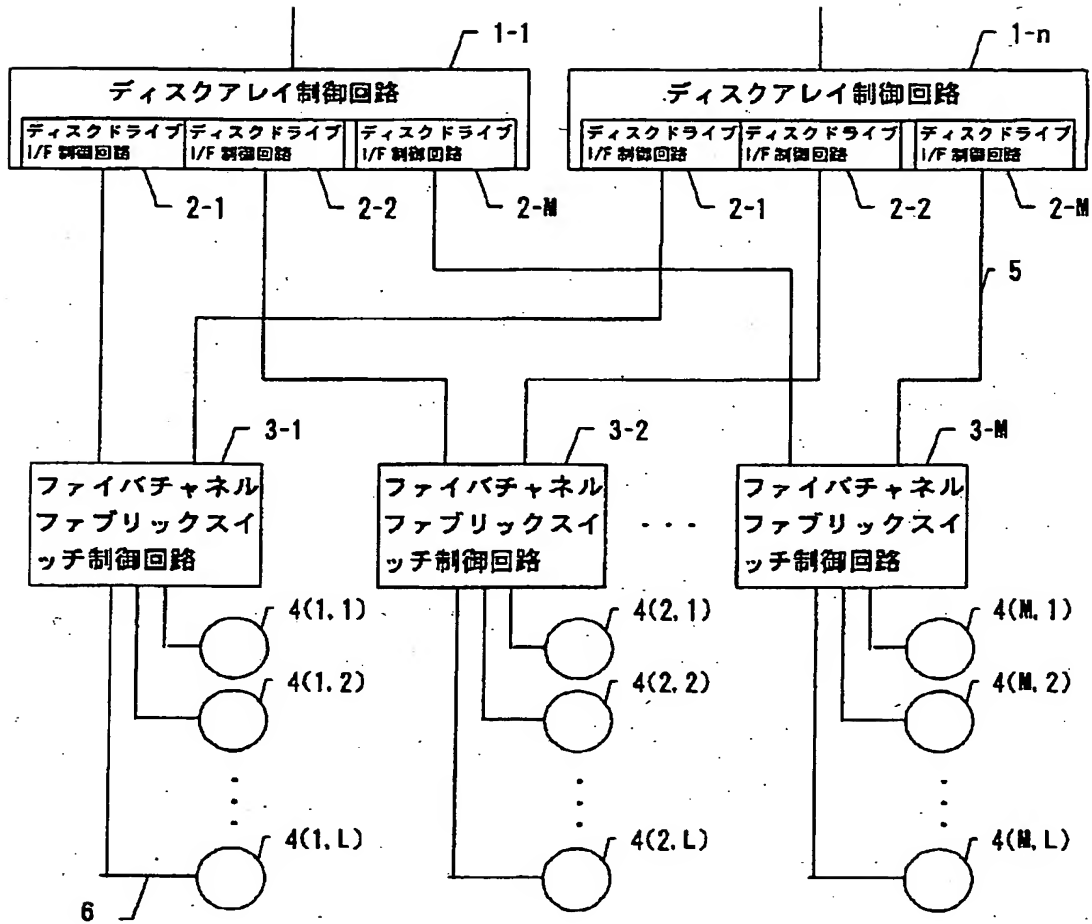
【図7】スペアディスク制御回路の接続図である。

【符号の説明】

1…ディスクアレイ制御回路、2…ディスクドライブインタフェース制御回路、3…ファイバチャネルファブリックスイッチ制御回路、4…ディスクドライブ、5…ファイバチャネルインタフェース、6…ファイバチャネルインタフェース、7…ホストインタフェース制御部、8…キャッシュメモリ、9…パリティデータ生成部、10…ディスクドライブグループ、11…ファイバチャネルアービトレイテッドループ制御回路、12…ディスクドライブ接続ポート、13…ループバイパス回路、14…ループバイパス信号切替信号、15…ファブリックスイッチ接続ポート、16…プロトコル制御部、17…スイッチ制御部、18…スイッチ、19…スペアディスク制御回路。

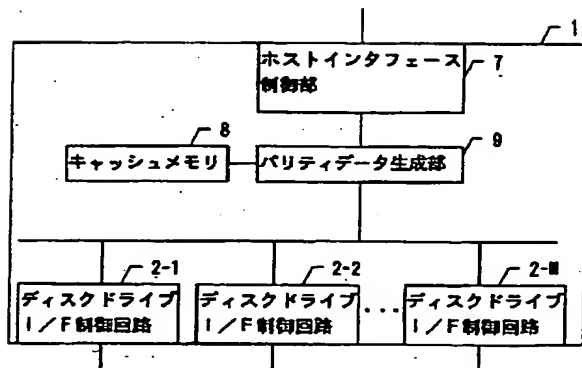
【図1】

図 1



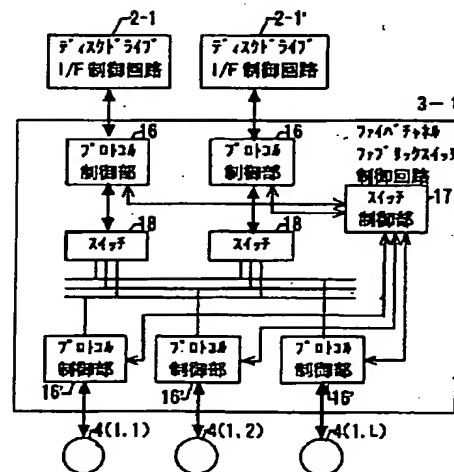
【図2】

図 2



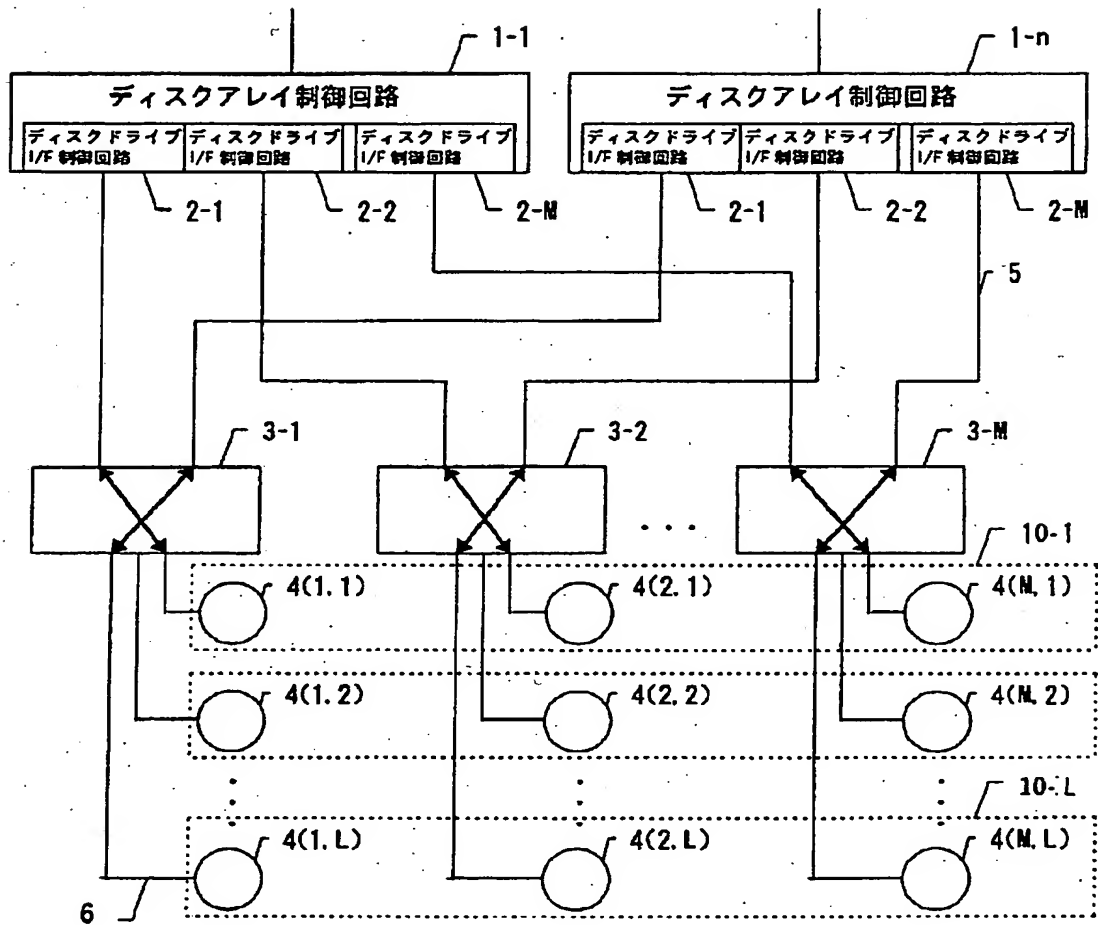
【図3】

図 3



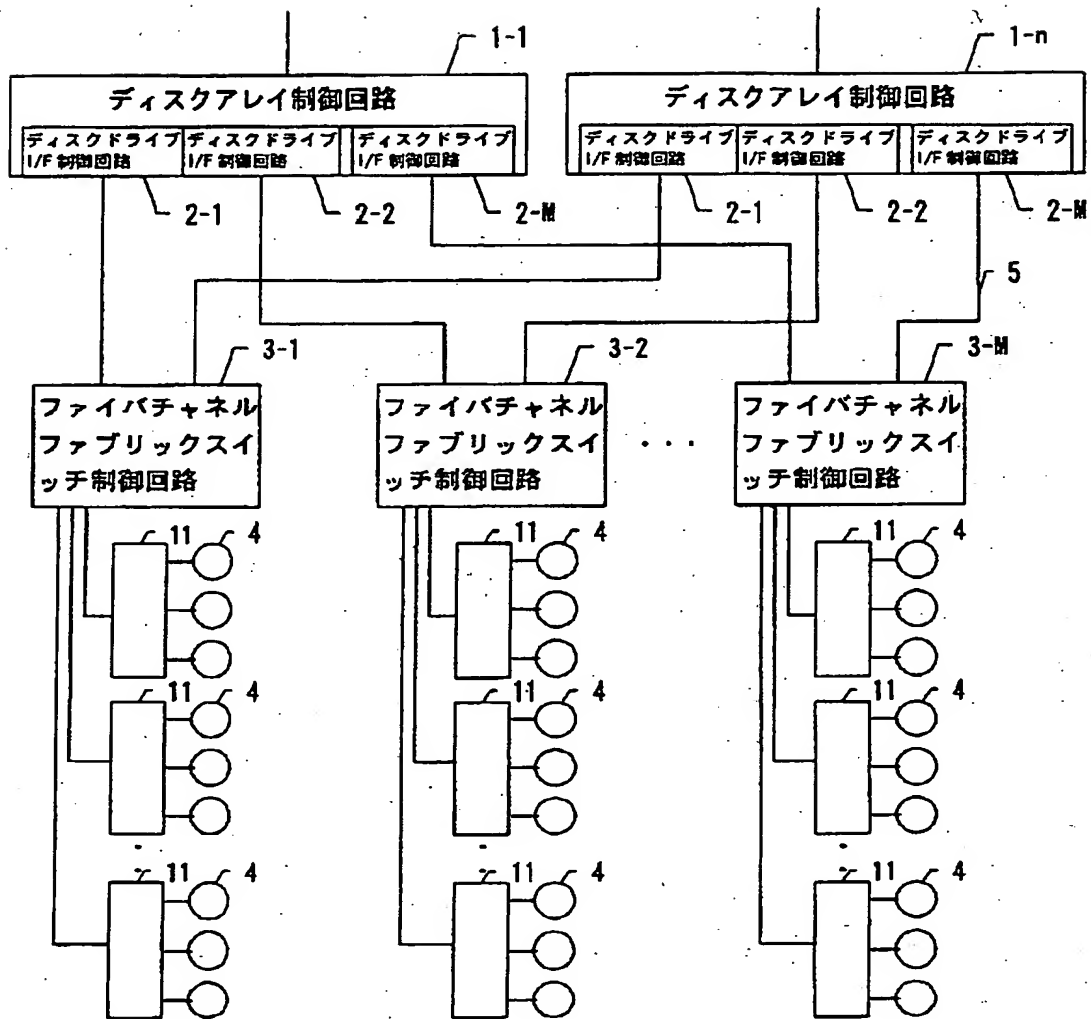
【図4】

図4



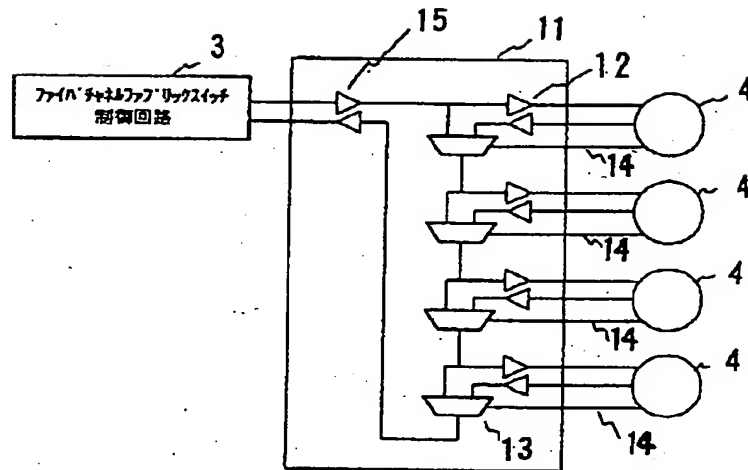
【図5】

図5



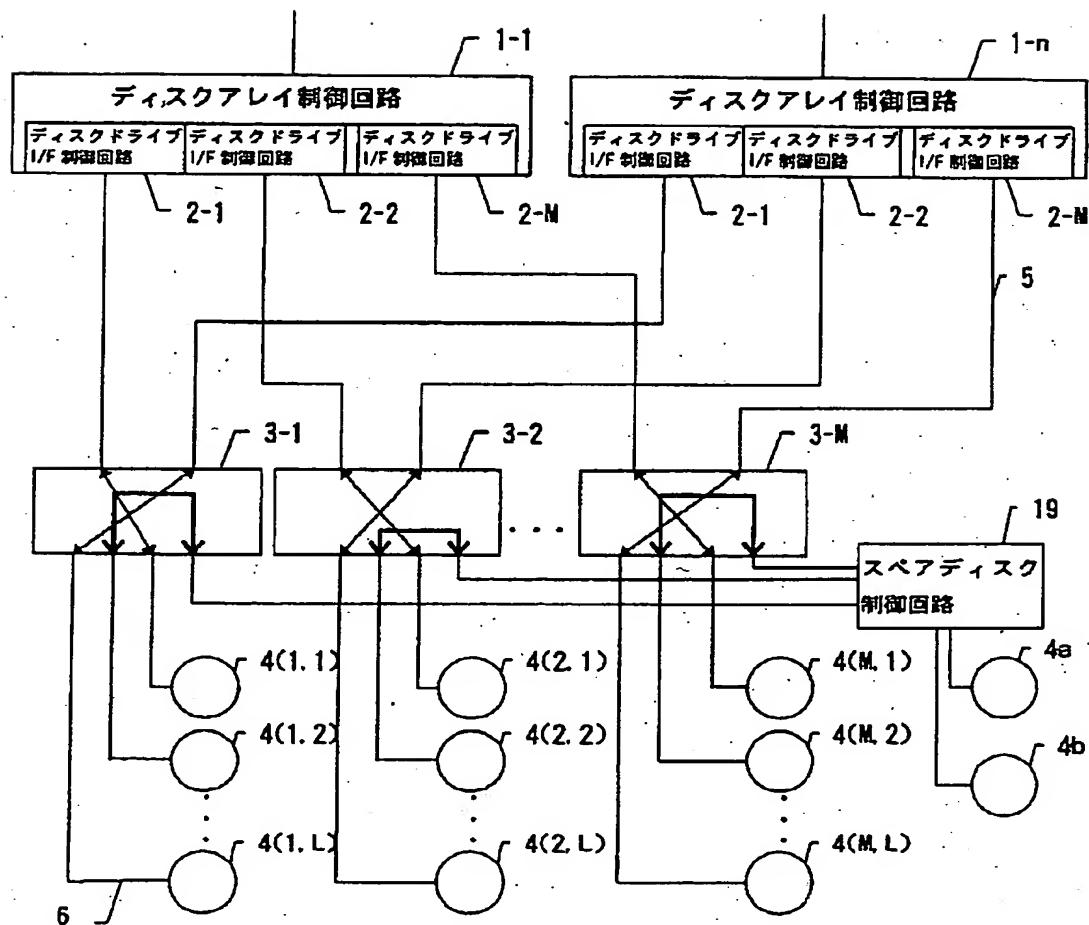
【図6】

図6



【図7】

図7



フロントページの続き

(51) Int. Cl.⁷

G 0 6 F 13/10

識別記号

3 4 0

F I

G 0 6 F 13/10

テマコード (参考)

3 4 0 A